

---

## Évaluer la maîtrise de la numération écrite chiffrée : choix du format QCM et validité d'items d'évaluations externes

*Assessment of student's knowledge about the writing of whole and decimal  
numbers: specificity of multiple-choice format and item's validity of external  
assessments*

**Nadine Grapin et Nathalie Sayac**

---



### Édition électronique

URL : <http://journals.openedition.org/educationdidactique/2836>

DOI : 10.4000/educationdidactique.2836

ISSN : 2111-4838

### Éditeur

Presses universitaires de Rennes

### Édition imprimée

Date de publication : 31 décembre 2017

Pagination : 55-72

ISBN : 978-2-7535-7319-2

ISSN : 1956-3485

### Référence électronique

Nadine Grapin et Nathalie Sayac, « Évaluer la maîtrise de la numération écrite chiffrée : choix du format QCM et validité d'items d'évaluations externes », *Éducation et didactique* [En ligne], 11-3 | 2017, mis en ligne le 31 décembre 2019, consulté le 04 janvier 2020. URL : <http://journals.openedition.org/educationdidactique/2836> ; DOI : 10.4000/educationdidactique.2836

---

# **ÉVALUER LA MAÎTRISE DE LA NUMÉRATION ÉCRITE CHIFFRÉE : CHOIX DU FORMAT QCM ET VALIDITÉ D'ITEMS D'ÉVALUATIONS EXTERNES**

Nadine Grapin

Laboratoire de didactique André Revuz (universités Paris Diderot, Paris-Est  
Créteil, Artois, Cergy Pontoise et Rouen). Université Paris-Diderot, ESPE de l'académie  
de Créteil

Nathalie Sayac

Laboratoire de didactique André Revuz (universités Paris Diderot, Paris-Est  
Créteil, Artois, Cergy Pontoise et Rouen). Université Paris-Diderot, ESPE de l'académie  
de Créteil

Si les QCM sont utilisés régulièrement dans les évaluations à grande échelle en mathématiques, la question de leur validité pour évaluer les connaissances des élèves de fin d'école en mathématiques a été encore peu abordée dans le cadre de la didactique des mathématiques. En comparant la réponse produite en QCM et celle fournie dans une question à réponse ouverte, puis en étudiant les stratégies mises en œuvre pour effectuer le choix de réponse, nous étudions la validité de la question en perspective de son format. Nous nous limitons ici à interroger le format QCM pour évaluer des connaissances relatives à l'écriture de nombres entiers et décimaux.

Mots-clés : évaluation externe, QCM, numération écrite chiffrée.

*Assessment of student's knowledge about the writing of whole and decimal numbers: specificity of multiple-choice format and item's validity of external assessments*

*If multiple-choice questions are frequently used in large scale mathematics assessments, few researches in didactic of mathematics have treated the question of the validity of such a format, especially concerning pupils' end of primary school (grade 5). In this paper, we analyze several tasks assessing knowledge of place value, including representing decimals and whole numbers using words, numbers or fractions. The validity of these tasks is studied by comparing answers given at a same task but with two different question formats (multiple choice and constructed response).*

*Keywords: external assessment, multiple-choice question, written positional notation.*

## INTRODUCTION

Le format de question QCM (Question à Choix Multiple) est utilisé régulièrement pour déterminer les connaissances des élèves de fin d'école primaire (10-11 ans, grade 5) en mathématiques dans les évaluations nationales ou internationales standardisées. En France, les élèves de cet âge et de ce niveau de scolarité ne sont pas habitués à être confrontés à des questions de ce type, que ce soit en situation d'évaluation ou non. Chercheuses en didactique des mathématiques nous avons mené plusieurs études autour du bilan CEDRE<sup>1</sup>2008 de fin d'école, en mathématiques (Sayac & Grapin, 2014a, 2015), puis il nous a paru également opportun d'étudier ce format spécifique de questions pour questionner la validité des évaluations nationales ou internationales qui l'emploient massivement.

Alors que l'évaluation nationale CEDRE 2014 conclut que seulement 57 % des élèves de fin d'école primaire ont « une connaissance solide des nombres entiers et une première connaissance stable des nombres décimaux » (Note d'information n° 18 – mai 2015), nous avons souhaité étudier sur quels « faits d'évaluation » (Chevallard & Feldman, 1986) s'appuie cette observation. La question qui nous intéresse dans cet article ne porte pas sur l'ensemble des tâches qui ont conduit à cette conclusion, mais plutôt sur les tâches qui nécessitent des conversions de registres (Duval, 1996) et sur le choix du format QCM pour évaluer ces connaissances ; nous nous focalisons uniquement sur l'écriture des nombres entiers et décimaux et sur des items portant sur des conversions « écriture en lettres – écriture chiffrée » pour les entiers et « fraction décimale – écriture à virgule » pour les décimaux. Le format QCM est-il pertinent pour évaluer de telles connaissances ? Comment se présentent les questions relatives à ce type de tâches ? Quelles sont les limites de ce format ? Les élèves mobilisent-ils effectivement les connaissances supposées être évaluées pour répondre à la question posée ?

Si ces questions rejoignent celles plus générales portant sur la validité des tests, il s'agit pour nous de les traiter dans un cadre didactique, et non pas dans celui de l'évaluation ou de la psychométrie comme c'est généralement le cas. En effet, les travaux issus de ces champs de recherche interrogent la qualité des QCM principalement à partir des caractéristiques statistiques des items (par exemple, Braibant, 2015)

et pointent certains dysfonctionnements, mais sans évoquer le contenu même de l'item, la tâche ou les connaissances mises en jeu pour le résoudre. Nous avons donc extrait des évaluations CEDRE 2008 et 2014 différents items proposés en format QCM et relatifs à des transformations d'écritures sur les nombres entiers et décimaux, puis nous les avons analysés afin d'étudier leur validité. Nous avons voulu observer, à travers un nombre d'exemples limité, la façon dont diffèrent les réponses produites par un même élève lorsque l'item est proposé sous un format ouvert et sous un format QCM et nous montrons comment une analyse *a priori* de la tâche s'avère nécessaire pour cibler les connaissances qui peuvent être évaluées, mais aussi pourquoi celle-ci n'est pas toujours suffisante pour garantir de la validité d'un item. En effet, l'étude des processus réellement mis en jeu par les élèves en situation de résolution de QCM témoigne de la mobilisation de stratégies qui ne reposent pas toutes sur des connaissances mathématiques et qui questionnent par conséquent ce que l'item permet réellement d'évaluer (Sayac & Grapin, 2014a, 2014b).

Nous présentons d'abord la façon dont la validité peut être abordée dans un cadre didactique afin de situer notre travail. Nous présentons ensuite l'étude que nous avons menée sur la validité d'items en format QCM portant sur des transformations d'écriture de nombres et concluons sur la pertinence de l'usage d'items sous la forme de QCM pour évaluer les connaissances des élèves et sur l'exploitation de tels items dans les évaluations.

## QUESTIONNEMENTS DIDACTIQUES SUR LA VALIDITÉ DES QCM EN ÉVALUATION

La question du format de question et de son influence sur le traitement de la tâche par l'élève a déjà été prise en compte dans différentes recherches en didactique des mathématiques (Adda, 1976 ; Duval & Pluvinage, 1977 ; Pluvinage, 1979 ; Maury, 1985 et plus récemment Vantourout & Maury, 2016) mais elles ne se situent pas forcément au niveau scolaire auquel nous nous plaçons (fin d'école). Si le format de question QCM a peu fait l'objet de recherches en didactique pour de jeunes élèves en France, il est davantage étudié dans le cadre d'évaluations à l'Université, notamment Choppin (1975), Leclercq (1986), Gilles (1996), mais aussi Tate

(2000), Forsyth et Spratt (1980) ou encore Lukhele, Thissen et Wainer (1994), Rodriguez (2003), Attali et Bar-Hillel (2003). En France, les QCM se développent en mathématiques depuis une dizaine d'années avec une utilisation désormais fréquente dans les épreuves du DNB (Diplôme National du Brevet) ou du Baccalauréat, mais aussi dans certains manuels de collège. Les QCM restent néanmoins encore absents des manuels de mathématiques utilisés dans les classes de l'école primaire, ce qui interroge la façon dont les élèves répondent à ce type de question et donc plus généralement la validité des items d'évaluation proposés avec ce format.

### Validité d'une épreuve : un point de vue didactique

Dans une conception partagée, un test est considéré comme valide s'il évalue bien ce qu'il est censé évaluer et uniquement cela ; plus précisément, c'est la validité des inférences qu'il est possible de faire à partir des scores des élèves qui est interrogée (Grégoire & Laveault, 2014). Si la validité des évaluations fait l'objet de nombreuses recherches en psychométrie, son étude d'un point de vue didactique est plus récente. Dans ce cadre, Grugeon et Grapin (2015) et Grapin (2015) ont développé une démarche permettant d'étudier la validité d'une évaluation en se référant, sur un domaine mathématique, à une organisation de mathématique de référence (Bosch & Gascon, 2005) ; l'étude peut alors être réalisée localement, sur chacun des items mais aussi, plus globalement, sur l'ensemble d'un domaine pour en étudier le recouvrement par les tâches sélectionnées dans l'évaluation, la représentativité de ces dernières, etc. Cette facette de la validité, qualifiée d'épistémo-didactique, repose alors sur l'analyse *a priori* de chacune des tâches pour étudier les potentialités de ces dernières et les savoirs mathématiques qui sont en jeu pour les résoudre.

De façon complémentaire, Vantourout et Goasdoué (2014) ont développé la notion de validité psycho-didactique d'un item : il s'agit, en particulier, d'étudier si « l'évalué va effectivement mobiliser la connaissance attendue » et de prendre en compte les processus et procédures mis en jeu par les élèves au regard de l'analyse *a priori*. Dans cette autre facette de la validité, au-delà de l'analyse *a priori* de la tâche, d'autres variables sont considérées telles que

le contexte de la situation, le format de la question, l'environnement (papier-crayon ou numérique), etc.

Les facettes épistémo et psycho-didactique de la validité se révèlent donc être complémentaires ; pour ce qui nous intéresse, la question de la validité se pose non seulement sur le contenu même du test, localement sur chacun des items, mais aussi sur l'adéquation entre les processus de réponses mis en jeu et ce que la tâche vise à évaluer. Dans le cas d'un item proposé classiquement sous une forme ouverte<sup>2</sup>(c'est-à-dire sans choix de réponse), l'analyse *a priori* menée dans un cadre didactique permet de déterminer, selon les valeurs des variables didactiques retenues pour la situation, une hiérarchie des stratégies que l'élève est susceptible de mettre en œuvre pour répondre et d'explicitier certaines procédures erronées.

Pour les QCM, les différents choix de réponses peuvent être considérés comme une variable didactique particulière, permettant des rétroactions, notamment en résolution de problèmes (Sayac & Grapin, 2014b), mais aussi, modifiant les processus de réponse mis en jeu par les élèves. Dans les tâches de conversions de registre qui nous intéressent ici, la différence de format des questions conduit, nous semble-t-il, à un changement important d'un point de vue cognitif. En effet, si en question ouverte, l'élève doit lui-même produire une représentation dans un registre donné en mettant en jeu une opération de conversion, il doit dans un QCM, reconnaître la représentation du nombre donné initialement parmi différentes représentations de ce même nombre, obtenues après conversion dans un autre registre. S'il semble alors nécessaire de proposer des items portant sur des tâches de conversion dans les deux sens, la distinction entre l'activité de conversion (en question ouverte) et celle de traitement reste posée en fonction du format de question.

### Conception et usage des QCM en situation d'évaluation

Dans les évaluations internationales, si on écarte les questions de type « vrai-faux » qui pourraient s'apparenter à un type spécifique de QCM, il est d'usage d'avoir trois distracteurs<sup>3</sup>et une seule réponse correcte. La conception du QCM repose en premier lieu sur la tâche elle-même, mais aussi sur la détermination des distracteurs. Il est d'usage que ces derniers

correspondent à des erreurs d'élèves (du moins, c'est ce qui est préconisé dans la conception de QCM [Leclercq, 1986, 2006]) ; l'analyse *a priori* de la tâche, menée après une étude épistémologique du savoir en jeu, permet bien souvent de fournir des distracteurs, comme nous le montrerons dans les paragraphes qui suivent. Néanmoins, il n'est pas toujours aisé d'en trouver trois différents qui soient pertinents.

Dans les évaluations standardisées CEDRE, le codage des réponses est binaire (correct ou incorrect) ; on calcule le score de l'élève en tenant compte uniquement des bonnes ou mauvaises réponses qu'il a fournies, sans interroger les différents types d'erreurs qu'il a pu faire. Il en est tout autrement dans les évaluations à visée diagnostique, comme par exemple les évaluations nationales qui ont été proposées en français et en mathématiques à tous les élèves en début de CE2 (grade 3), puis en début de sixième (grade 6) de 1989 à 2008. Dans ce cas, les réponses des élèves sont codées en tenant compte des procédures mises en jeu et le type d'erreurs potentielles ; pour les QCM, il est donc important que les distracteurs soient « bien choisis » et correspondent *a priori* à un certain type de fonctionnement cognitif erroné. La conception de l'évaluation Pépite en algèbre (Grugéon, 1997 ; Grugéon, Chenevotot-Quentin, Pilet & Delozanne, 2012) repose sur ce principe, les items s'appuyant sur une analyse épistémologique et didactique des savoirs en jeu, et les distracteurs correspondant à des erreurs spécifiques. Dans ce dispositif, les procédures et les justifications de réponses aux questions ouvertes ainsi que les choix de réponses des élèves aux QCM sont mis en rapport les uns avec les autres. Ils fournissent alors un faisceau d'indicateurs qui permet de déterminer des cohérences de fonctionnement et conduit à déterminer des profils d'élèves selon leurs réponses ; ces profils étant ensuite exploités pour mettre en place des parcours d'enseignement différencié (Pilet, 2012).

Quelle que soit la façon dont sont déterminés les distracteurs et la façon dont ils sont exploités par la suite, la question de l'impact de leur présence sur l'activité de l'élève reste posée, en particulier à travers les stratégies mises en œuvre pour répondre au QCM.

## Stratégies des élèves

La question des stratégies que développent les élèves lorsqu'ils sont confrontés à des QCM a été étudiée par de nombreux chercheurs (Choppin, 1975 ; Leclercq, 1986, 1987), mais elle a toujours été traitée pour des étudiants adultes et d'autre part, elle a toujours été étudiée du point de vue de la psychologie cognitive. Choppin (1975) a retenu 3 modèles<sup>4</sup>, combinant de manière variée connaissances des élèves et opportunités liées à la forme spécifique d'un QCM alors que Katz, Bennet, et Berger (2000) ont distingué les stratégies « *traditional* » des « *non traditional* ». Les premières reposant sur des procédures qui sont enseignées et que l'élève utiliserait pour répondre si la question était posée sous forme ouverte alors que les autres, spécifiquement associées à des QCM, sont celles où l'élève s'appuie sur les réponses proposées pour faire un choix.

De notre côté, nous avons souhaité prendre en compte les stratégies des élèves en prenant davantage en compte la façon dont ils convoquaient (ou non) leurs connaissances, même si les QCM ou QROC ne peuvent fournir qu'un nombre restreint d'informations relatives à ces connaissances (Joffe, 1992 ; Ginsburg *et al.*, 1992 ; Yackel, Cobb & Wood 1992). Dans le cadre d'une recherche précédente (Sayac & Grapin, 2014a), nous avons dans un premier temps listé, de manière aussi exhaustive que possible (Annexe 1), les différentes stratégies que pouvaient utiliser des élèves de fin d'école pour répondre à des QCM, puis nous les avons regroupées suivant trois types selon qu'elles intégraient des niveaux différents de connaissances :

- stratégies A : *stratégies de savoirs*. L'élève active des connaissances ou des savoir-faire (techniques – raisonnement) pour choisir la réponse qu'il pense être la bonne : soit il résout complètement la tâche (par la procédure de son choix, juste ou fausse) : ce qui s'apparente aux stratégies « *traditionnelles* » définies par Katz, Bennet et Berger (2000), soit il teste les propositions de réponse et choisit celle qui peut convenir ;
- stratégies B : *stratégies de substitution ou de repli*. L'élève n'utilise pas ses connaissances mathématiques de façon explicite pour faire un choix : son choix ne repose pas de façon assurée sur ses connaissances, mais sur des pseudo-connaissances ou « *intuitions cognitives* ». Parmi ces stratégies, figure, entre

autres, celle de répondre au hasard ou encore de reprendre les nombres en présence dans l'énoncé en les combinant éventuellement, mais sans réel raisonnement mathématique pour trouver un des choix de réponses proposés ;

– stratégies C : *stratégies mixtes*. L'élève a initié un raisonnement pour répondre à la question posée, mais il se sert des différentes propositions de réponse pour finaliser son choix.

La recherche que nous avons menée (Sayac & Grapin, 2014a) à partir d'une expérimentation intégrant des QCM nous avait permis de constater que les élèves mobilisaient des stratégies de réponse variées pour un même test et qu'ils étaient capables d'en changer en fonction de leur niveau de connaissance, ce qui interrogeait la validité psycho-didactique de certains items. Nous avons choisi d'adjoindre des degrés de certitude<sup>5</sup> aux réponses fournies par les élèves, ce qui nous a amenées à réaliser que les élèves les plus faibles ne se déclaraient pas en manque d'assurance, surtout pour les items portant sur les conversions d'écriture de nombres décimaux (passage d'une écriture sous forme de fraction décimale en une écriture à virgule et réciproquement).

Nous avons voulu compléter cette première recherche sur les stratégies de réponse par un questionnement plus ciblé sur la différence de format de question qui a déjà étudié notamment par Bennett, Rock et Wang (1991), Frary (1985) ou Bridgeman (1992). Dans cette seconde recherche, nous avons souhaité évaluer l'impact du format de la question sur la réussite et les stratégies des élèves et nous nous sommes posées les questions suivantes : est-ce que, pour des items portant sur des transformations d'écriture, le format de la question impacte sur les processus mis en jeu par l'élève pour répondre et de quelle manière ? En particulier, est-ce qu'une question sous forme de QCM garde le même potentiel évaluatif qu'une question sous forme ouverte, c'est-à-dire, est-ce que l'activité de l'élève est « conforme » à celle prévue par le concepteur et permet de révéler les mêmes connaissances qu'une question sous forme ouverte ? Qu'en est-il alors de la validité psycho-didactique d'items construits sous la forme de QCM et apparaissant dans les évaluations nationales ?

## ÉTUDE DIDACTIQUE DE LA VALIDITÉ D'ITEMS EN FORMAT QCM : COMPARAISON AVEC UN FORMAT OUVERT ET STRATÉGIES DE RÉOLUTION

Pour étudier d'un point de vue didactique la validité d'items d'évaluation, nous avons voulu observer à la fois les stratégies mises en œuvre par les élèves pour répondre à des items présentant un format spécifique (QCM), mais aussi étudier les différences de réponses contingentes aux différents formats de question (ouvert ou QCM).

Nous avons donc réalisé une expérimentation confrontant deux tests équivalents du point de vue des tâches mathématiques, mais différents par leur format : l'un avec des QCM extraits de CEDRE 2008 ou 2014, l'autre avec des Questions à Réponse Ouverte Courte (QROC). Nous avons réalisé cette expérimentation en juin 2013 auprès de 195 élèves de CM2 (grade 5), de 9 classes différentes, qui ont été soumis successivement à deux tests, l'un exclusivement sous format de QCM, l'autre reprenant les mêmes tâches mathématiques, mais présentées sous format QROC.

Neuf items de transformation d'écriture ont été passés dans le cadre de cette expérimentation : deux QCM sur les nombres entiers et deux QCM sur les nombres décimaux, trois QROC sur les nombres entiers et deux QROC sur les décimaux.

L'analyse *a priori* de ces différentes tâches a permis de déterminer les savoirs mathématiques qui étaient en jeu et les connaissances qui pouvaient être évaluées par ces items. Pour analyser les stratégies mises en œuvre par les élèves pour répondre au QCM, nous avons observé individuellement et successivement chacun des 195 élèves passant le premier test (sous forme de QCM) et avons éventuellement demandé à l'élève de nous préciser oralement la façon dont il avait procédé pour faire son choix de réponse. Nous avons ainsi pu coder, pour chaque élève, la stratégie qu'il avait adoptée pour répondre aux QCM. Le test passé sous forme de QROC a été réalisé en classe, généralement la veille du second test pour qu'entre les deux tests, aucun enseignement sur l'écriture des nombres ait pu avoir lieu et sans accompagnement spécifique ; l'ensemble des réponses aux deux tests a été codé pour pouvoir comparer les différences de réponses produites selon le format de question.



## ÉTUDE DES NOMBRES ENTIERS : (A-) PERTINENCE DU FORMAT QCM

### Écriture des nombres entiers : construction de QCM et stratégies

Convertir une écriture en lettres en une écriture chiffrée est un des objectifs usuels des tâches d'évaluation de fin d'école sur l'écriture des nombres entiers. La non congruence entre ces deux registres de représentation a déjà été soulignée à de nombreuses reprises (par exemple, Mounier, 2010) et est source de difficulté, en particulier au début de l'apprentissage de l'écriture des nombres entiers; une erreur classique étant justement de mobiliser des principes de congruence entre la numération écrite chiffrée et la numération parlée, et par exemple d'écrire « quatre-cent-huit », 4008 ou 4 100 8. Évaluer cette connaissance par un QCM, au début de l'apprentissage, demande donc de proposer des distracteurs qui reposent sur ce type d'erreur, mais est-ce toujours pertinent en fin d'apprentissage ?

Pour évaluer les connaissances relatives au passage d'une écriture en lettres en une écriture chiffrée, la question ne porte pas toujours sur l'écriture effective du nombre, comme le montre l'item QCM 2 suivant (Tableau 1).

Il s'agit ici de donner le nombre de chiffres nécessaires pour écrire le nombre. Pour ce faire, l'élève peut soit écrire le nombre puis compter le nombre de chiffres qui le composent soit mettre en relation l'écriture (ou la lecture) du nombre avec les classes qui le composent et le nombre de chiffres correspondant.

Dans cette question, les nombres en jeu sont de l'ordre des millions ou de la centaine de mille (l'écriture chiffrée comporte entre 6 et 9 chiffres); ce choix relatif à la taille des nombres est en cohérence avec les programmes de cycle 3 actuellement en vigueur. Chacun des choix de réponses proposés (6 – 7 – 8 et 9 chiffres) correspond pour chacun d'eux à une bonne réponse et une seule : il n'y a pas de proposition de réponse qui serait correcte deux fois ou qui ne le serait jamais.

Figure 1.  
Énoncé du QCM 1

QCM 1 : Deux-millions-deux-cent-vingt-cinq-mille-six s'écrit :			
<input type="checkbox"/> 2 225 600	<input type="checkbox"/> 2 522 006	<input type="checkbox"/> 2 225 006	<input type="checkbox"/> 2 225 060

Le premier item (QCM1, FIG. 1) que nous avons extrait de l'évaluation CEDRE 2008 vise à associer écriture littérale et écriture chiffrée d'un grand nombre. Le découpage par classes est l'objet principal de cet item. La difficulté principale se situe sur la classe des unités simples et sur la reconnaissance de l'absence des unités de centaine et dizaine qui se traduit par la présence de deux zéros dans l'écriture chiffrée du nombre. Les distracteurs proposés ne sont pas construits à partir de la non-congruence évoquée ci-dessus, mais ils sont obtenus par une permutation des chiffres composant le nombre.

La présentation sous forme de tableau de plusieurs QCM enchaînés, pas forcément usuelle pour les élèves, permet néanmoins deux types de rétroactions : amener l'élève à éliminer des réponses qui ne seraient pas 6 – 7 – 8 ou 9 et à éventuellement contrôler ses réponses en s'assurant qu'il a mis une seule croix dans chacune des colonnes (ce qui n'est pas une stratégie de contrôle assurée mathématiquement, mais permise par les choix de réponse proposés). En revanche, dans cet item, comme pour le précédent, le choix d'un distracteur plutôt qu'un autre ne pourra pas indiquer le type d'erreur que

Tableau 1. Énoncé du QCM 2

QCM 2	6 chiffres	7 chiffres	8 chiffres	9 chiffres
QCM 2a Sept-millions-cinq-cent-cinq-mille	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QCM 2b Vingt-cinq-millions-cinq-cent-mille	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QCM 2c Cinq-cent-mille	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QCM 2d Sept-cent-millions-cinq-cent-mille	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

produit l'élève et par conséquent, il nous semble particulièrement intéressant d'étudier la façon dont les élèves ont procédé pour répondre à cet item.

Dans l'expérimentation que nous avons menée, l'item QCM 1 a été réussi par 88 % des élèves, et l'item QCM 2 a été réussi (4 réponses justes sur les 4) par 62 % des mêmes élèves ; la répartition des réponses des élèves selon les choix de réponse figure en annexe 2. Comment expliquer cet écart de score entre les deux questions ? Nous avons observé que pour répondre à ces deux items, les élèves ont très majoritairement employé des stratégies de savoirs (Annexe 3) : pour le QCM 1, 96 % des élèves ont procédé en regardant chiffre à chiffre ou classe par classe chacune des propositions ou en écrivant le nombre et en le comparant ensuite avec ceux proposés dans le QCM. Ils sont 92 % à avoir utilisé ces mêmes stratégies pour le QCM 2, en comptant mentalement le nombre de chiffres, ou en écrivant d'abord le nombre puis en comptant les chiffres ensuite, ou encore en procédant classe par classe. L'écart de scores peut s'expliquer par le fait que dans l'item QCM 2, le même type de connaissance est évalué 4 fois de suite et par conséquent, la possibilité de faire une erreur augmente (14 % des élèves n'ont d'ailleurs fait qu'une seule erreur sur leurs 4 réponses).

### QCM vs QROC : quel format pour quelle évaluation ?

Nous avons comparé les réponses du QCM 1 avec celles du QROC 1a suivant équivalent du point de vue de l'analyse a priori de la tâche mathématique à réaliser :

**QROC 1a :** écrire en chiffres « cinq-millions-cinq-cent-cinquante-deux-mille-six ».

Si la réussite au QCM 1 est de 88 %, elle n'est que de 70 % au QROC 1a, mais certaines réponses erronées au QROC semblent relever d'un manque d'attention plutôt que d'une méconnaissance majeure, notamment pour l'écriture d'un chiffre à la place d'un autre, comme par exemple 5 542 006 ou lieu

de 5 552 006<sup>6</sup>. Même si ce sont des stratégies de savoir qui sont employées pour répondre au QCM (Annexe 3), elles ne mettent pas en jeu les mêmes procédures qu'en QROC : si l'élève doit écrire le nombre en QROC, dans le QCM il doit reconnaître une écriture. Par ailleurs, nous avons constaté qu'aucune erreur produite à la question ouverte ne correspond aux types de choix proposés dans le QCM, ce qui remet en cause le choix des distracteurs proposés et peut aussi expliquer l'écart de réussite de plus de 10 % entre les deux questions. À ce titre, rétrospectivement, nous avons constaté une grande variété des erreurs produites en ouvert (Annexe 2) et, même si la moitié de celles-ci reposent sur la méconnaissance des règles de la numération écrite chiffrée en lien avec la transcription orale du nom du nombre, les productions erronées des élèves restent très diversifiées. Ce constat rend difficile la transformation d'une QROC en QCM, notamment pour la détermination des distracteurs à partir de réponses d'élèves ; les choisir à partir de l'analyse didactique serait un gage de validité supplémentaire, mais pour autant, ils ne correspondraient pas nécessairement à une des erreurs faites par les élèves en question ouverte.

Nous avons également constaté qu'un échec au QCM n'impliquait pas nécessairement un échec en QROC, et réciproquement. En effet, parmi les 23 élèves qui ont coché un des distracteurs au QCM 1, seuls 11 se sont également trompés au QROC 1a. Réciproquement, sur les 9 élèves qui avaient fait une erreur que nous avons qualifiée d'inattention à la QROC 1a, 7 ont coché la bonne réponse au QCM 1, ce qui conforte notre analyse. Même si l'échantillon est faible et que nous ciblons une seule tâche, la question de l'impact de la différence de format sur l'activité de l'élève semble donc une piste intéressante pour mieux comprendre le fonctionnement cognitif des élèves en situation d'évaluation.

Pour essayer d'aller plus loin dans l'évaluation des connaissances des élèves de fin d'école au niveau de l'écriture des nombres entiers, nous avons choisi de comparer les réponses données au QCM2 avec deux tâches QROC différentes (FIG. 2), mais connexes.

Figure 2.

Énoncé des QROC 1b et 2

**QROC 1b :** écrire en chiffres : huit-millions-six-cent-six-mille ;  
**QROC 2 :** quel est le nombre de chiffres nécessaires pour écrire chacun des nombres ?



Tableau 2.  
Énoncé de la QROC 2

	Nombre de chiffres
QROC 2a : Trente-quatre-millions-cinq-cent-mille	.....
QROC 2b : Neuf-cent-mille	.....
QROC 2c : Huit-cent-millions-deux-cent-mille	.....

Les écritures chiffrées des nombres impliqués dans les QCM 2 et QROC 1b et 2 (Tableau 2) ont des tailles identiques et se terminent toutes par cinq zéros ; seuls les autres chiffres diffèrent. Par conséquent, nous pouvons considérer que la QROC 2 est équivalente du point de vue de l'analyse *a priori*, à la tâche à réaliser au QCM 2. Nous avons également voulu savoir ce que permettait d'évaluer ce type de tâche en le comparant directement avec une question ouverte (QROC 1b) où l'élève doit écrire le nombre demandé ; ce qui est en lien avec une des procédures possibles pour résoudre, à savoir écrire le nombre en chiffres et compter le nombre de chiffres.

Comme pour le premier item, nous constatons une meilleure réussite au QCM, mais avec des écarts de moindre importance avec le format ouvert, puisqu'ils varient selon les nombres de 5 % à 16 % (Tableau 3).

Tableau 3.  
Scores de réussite selon le format de question

	QCM	QROC	Écart QCM – QROC
QCM 2a – QROC 1b	78 %	70 %	8 %
QCM 2b – QROC 2a	74 %	69 %	5 %
QCM 2c – QROC 2b	92 %	76 %	16 %
QCM 2d – QROC 2c	77 %	70 %	7 %

Nous expliquons ces écarts à la fois par les potentiels effets rétroactifs du QCM et par une meilleure compréhension de la consigne sous format QCM (9 élèves, soit 4,6 % de l'échantillon, dans la QROC 2, ont écrit le nombre demandé au lieu du nombre de chiffres qui composent son écriture).

Seul l'écart de score le plus important (16 %) est significatif : il correspond à l'écriture du seul nombre de l'ordre de la centaine de mille (et non du million). Les élèves étant plus familiers avec des nombres de

cet ordre réussissent mieux en ouvert, mais la réussite est très importante en QCM. Ce score particulièrement élevé (92 %) peut s'expliquer de différentes façons : d'une part, comme c'est le seul nombre qui n'est pas de l'ordre du million, l'élève peut choisir le nombre de chiffres le plus petit et trouver la bonne réponse puisque tous les choix de réponse sont corrects pour une des questions. D'autre part, si l'élève associe (avec une stratégie de savoir incorrecte) la longueur du nom du nombre avec l'écriture en chiffres, il est encore une fois amené à choisir le nombre de chiffres le plus petit et à trouver la bonne réponse.

Si nous comparons la cohérence des réponses entre l'écriture du nombre sous format ouvert (QROC 1b) et le nombre de chiffres nécessaires sous forme de QCM (QCM 2a), nous constatons que parmi les 58 élèves qui se trompent dans l'écriture du nombre, la moitié (31 élèves pour environ 5 % des élèves) réussit à donner le bon nombre de chiffres au QCM. Parmi ces 31 élèves, 13 d'entre eux écrivaient déjà le nombre en ouvert avec 7 chiffres, mais de façon erronée, avec des zéros mal positionnés. Au-delà de la comparaison des scores de réussite QROC-QCM, c'est la tâche elle-même que nous interrogeons ici : demander uniquement le nombre de chiffres pour évaluer la connaissance relative à l'écriture d'un nombre ne semble ni suffisante, ni pertinente puisqu'elle ne permet ni de repérer les erreurs produites par les élèves ni d'évaluer la connaissance visée (des stratégies autres que celles de savoir pouvant intervenir pour produire la réponse correcte). Ce type de tâche ne nous semble donc pas qualifié pour évaluer de manière valide les connaissances des élèves relatives à l'écriture des nombres.

## ÉCRITURE DE NOMBRES DÉCIMAUX : (A)-PERTINENCE DU FORMAT QCM

Si les difficultés inhérentes aux aspects décimal et positionnel de la numération écrite chiffrée ont été soulevées à de nombreuses reprises pour les nombres entiers, il en est de même pour celles relatives à la conversion d'écritures entre le registre des fractions décimales et celui des écritures à virgule (Comiti & Neyret, 1979 ; Douady & Perrin-Glorian, 1986 ; Roditi, 2007 ; Chesné, 2014). Nous nous limitons ici à cet aspect des nombres décimaux, mais il est évident que l'évaluation seule de ces conversions est

insuffisante pour cerner les connaissances des élèves sur l'écriture des décimaux.

### Écriture des nombres décimaux : construction de QCM et stratégies

À la différence des items sur les nombres entiers, les distracteurs des items que nous avons retenus (QCM 3 et QCM 4) correspondent à des erreurs repérées dans les différents travaux de didactique chez les élèves de ce niveau scolaire ; nous les explicitions à partir des deux QCM que nous avons proposés (Tableaux 4 et 5).

Tableau 4.

Explication des choix des distracteurs pour le QCM 3

QCM 3 : À quel nombre correspond la fraction $\frac{62}{10}$ ?	
<input type="checkbox"/> 6,2	bonne réponse
<input type="checkbox"/> 0,62	confusion dixième – centième
<input type="checkbox"/> 62,10	conception erronée : $a/b = a,b$
<input type="checkbox"/> 620	confusion « diviser par 10 » et « multiplier par 10 »

Tableau 5.

Explication des choix des distracteurs pour le QCM 4

QCM 4 : Quelle est la fraction égale à 237,8 ?	
<input type="checkbox"/> $\frac{237}{8}$	conception erronée : $a/b = a,b$
<input type="checkbox"/> $\frac{2378}{100}$	confusion dixième – centième
<input type="checkbox"/> $\frac{2378}{10}$	bonne réponse
<input type="checkbox"/> $\frac{237}{10}$	prise en compte uniquement de la partie entière (237)

Le QCM 3 est représentatif d'une conversion entre le registre des écritures en fractions décimales et celui des écritures à virgule (sens  $Q \rightarrow D$ ) alors que le QCM 4 est représentatif de la conversion inverse (sens  $D \rightarrow Q$ ). Nous constatons que les choix des élèves se répartissent différemment sur les deux QCM (Tableaux 6 et 7) : le score de réussite est plus élevé au QCM 3, mais l'erreur consistant à consi-

dérer  $a,b$  comme étant une écriture de  $a/b$  est plus fréquente dans le QCM 4 (20 %) que dans le QCM 3 (13,8 %). Il est difficile d'en déterminer la cause, mais le fait que, dans l'évaluation CEDRE, cette mauvaise réponse soit donnée en premier choix dans le QCM 4 a pu induire davantage les élèves en erreur.

Tableau 6.

Répartition des choix de réponses des élèves selon les propositions de réponse au QCM 3

QCM 3 : À quel nombre $\frac{62}{10}$ correspond la fraction $\frac{62}{10}$ ?		Nombre d'élèves	Pourcentage
<input type="checkbox"/> 6,2	bonne réponse	147	75,4 %
<input type="checkbox"/> 0,62	confusion dixième – centième	17	8,7 %
<input type="checkbox"/> 62,10	conception erronée : $a/b = a,b$	27	13,8 %
<input type="checkbox"/> 620	confusion « diviser par 10 » – « multiplier par 10 »	4	2,1 %

Tableau 7.

Répartition des choix de réponses des élèves selon les propositions de réponse au QCM 4

QCM 4 : Écrire sous forme de fraction 237,8		Nombre d'élèves	Pourcentage
<input type="checkbox"/> $\frac{237}{8}$	conception erronée : $a/b = a,b$	39	20,0 %
<input type="checkbox"/> $\frac{2378}{100}$	confusion dixième – centième	12	6,2 %
<input type="checkbox"/> $\frac{2378}{10}$	bonne réponse	137	70,3 %
<input type="checkbox"/> $\frac{237}{10}$	prise en compte uniquement de la partie entière	7	3,6 %

Si nous observons des écarts sur les choix de réponse, qu'en est-il des stratégies mises en œuvre par les élèves pour répondre à ces QCM ? Pour le QCM 3 comme pour le QCM 4, ce sont bien des stratégies de savoir<sup>7</sup> qui sont employées (environ 90 % des élèves pour chacun des QCM), mais cela ne signifie pas pour autant que les procédures ou les règles utilisées par les élèves pour gérer ces conversions sont correctes :

- pour le QCM 3, parmi les 178 élèves qui emploient des stratégies de savoir, seuls 78 % d'entre eux choisissent ensuite la bonne réponse ;
- pour le QCM 4, parmi les 175 élèves ayant mis en œuvre une stratégie de savoir, ils sont 72 % à répondre correctement.

Cela signifie donc que les élèves utilisent des théorèmes en acte (Vergnaud, 1981) erronés pour répondre à ces questions.

### QCM vs QROC : quel format pour quelle évaluation ?

Pour affiner notre analyse, nous allons à présent confronter les résultats des élèves selon le format de questions proposé. Les questions ouvertes correspondant aux QCM 3 et 4 sont les suivantes (FIG. 3) :

Figure 3.  
Énoncé des QROC 3 et 4

QROC 3 : Donner une écriture à virgule de la fraction $\frac{95}{10}$
QROC 4 : Écrire 149,7 sous forme d'une fraction

Tableau 8.  
Répartition des choix de réponses des élèves selon les propositions de réponse au QCM 4

		QROC	QCM
Passage d'une écriture fractionnaire à une écriture à virgule (Q → D) QROC 3 et QCM 3	Réussite	153 élèves 78,5 %	143 élèves 75,4 %
	Non réponse	5 élèves 2,6 %	0
Passage d'une écriture à virgule à une écriture fractionnaire (D → Q) QROC 4 et QCM 4	Réussite	101 élèves 51,8 %	137 élèves 70,3 %
	Non réponse	21 élèves 10,8 %	0

Comme le montrent les résultats du tableau 8, la différence de format (QROC – QCM) a un impact variable sur la réussite des élèves selon le sens de la transformation en jeu : dans le sens Q → D (QROC et QCM 3), les taux de réussite des élèves passent

de 75,4 % (QCM 3) à 78,5 % (QROC 3) alors que dans le sens inverse (D → Q, QROC et QCM 4), la différence de réussite est ici significative en faveur du QCM (70,3 % pour le QCM 4 et 51,8 % pour le QROC 4). Cette différence importante de réussite, selon le format, dans le sens Q → D, s'explique en partie par un score important de non réponse en ouvert (environ 11 %) et par un effet rétroactif du QCM sur certaines réponses, qui ont été produites par les élèves en ouvert, mais que les élèves ne retrouvent pas dans les choix (en particulier, des écritures correctes, du type  $237 + 8/10$  mais qui ne répondent pas complètement à la question posée, puisqu'elles ne sont pas sous forme de fraction). Ce résultat vient ainsi illustrer le fait qu'une conversion de registre est « orientée et non réversible. Ce qui veut dire que la conversion directe et la conversion inverse sont souvent de nature différente et soulèvent des difficultés et des obstacles qui n'ont rien de commun. » (Duval, 2006). Dans notre cas, un élève de fin d'école primaire associe généralement une seule écriture à virgule à la fraction décimale  $62/10$ , alors qu'il peut associer différentes écritures fractionnaires à un nombre en écriture à virgule :  $237,8 = 2378/100 = 23780/1000 = 237 + 8/10...$

Comme nous l'avons déjà précisé, la différence de format implique un changement dans l'activité de conversion de l'élève (conversion dans le format QROC et reconnaissance dans le format QCM) ; voyons plus précisément comment cela influence la réponse donnée par les élèves dans chacun des formats.

Dans le sens Q → D (QROC 3) : sur les 153 élèves qui répondent correctement en QROC, 88 % répondent aussi correctement au QCM. Sur les 12 élèves qui répondent que  $95/10 = 0,95$  en QROC, la moitié d'entre eux font le même type d'erreur en QCM. Enfin, parmi les 13 élèves qui répondent 95,10 en QROC, 10 produisent une réponse similaire au QCM. Au-delà des pourcentages de réussites proches entre la QROC et le QCM (écart de 4 %), nous constatons que les réponses des élèves ne sont pas forcément « cohérentes » entre la QROC et la QCM. En particulier, 21 élèves, soit 11 % de notre échantillon, répondent correctement en QROC et se trompent au QCM.

Si nous menons une analyse similaire pour le sens D → Q (QROC 4) : sur les 101 élèves qui écrivent que 149,7 =  $1497/10$ , 91 % d'entre eux répondent également correctement au QCM. Sur les 16 élèves qui

répondent 149/7 en QROC, 10 répondent 237/8 et les 6 autres choisissent la bonne réponse dans le QCM. Parmi les 16 élèves qui répondent 1 497/1 000 en QROC, seuls 4 élèves réitèrent leur erreur au QCM, mais les 12 autres choisissent la bonne réponse. Seuls 2 élèves sur les 7 qui ont écrit que  $149,7 = 149/10$  font le même choix en QCM. Enfin, deux élèves seulement ont donné la réponse 1 497/100 en QROC mais ne l'ont pas choisi au QCM.

Nous constatons donc que pour ces deux tâches de conversion, la cohérence dans les réponses proposées est variable et que le format de la question impacte sur la réponse de l'élève. Dans le sens  $D \rightarrow Q$ , nous observons que le QCM ne favorise pas systématiquement la réussite puisque 11 % des élèves réussissent en ouvert et se trompent au QCM. Il n'en est pas de même dans le sens  $Q \rightarrow D$ , puisqu'ici les élèves qui se trompent en ouvert, modifient pour beaucoup leur réponse pour choisir la bonne réponse en QCM. Ce premier constat rejoint le fait que l'utilisation de QCM en cours d'apprentissage n'est pas préconisée (Grégoire & Laveault, 2014), les connaissances des élèves n'étant pas encore suffisamment « stabilisées » ; il nous conduit aussi à penser que pour évaluer les connaissances des élèves et rechercher des cohérences de fonctionnement, l'étude des réponses à des items pris isolément ne peut suffire ; comme le fait Grugeon (1997), la mise en perspective des réponses à un nombre important d'items se révèle nécessaire pour rechercher ces cohérences, mais est-elle réellement suffisante ?

## CONCLUSION

En ce qui concerne les QCM présents dans cette étude, nous avons constaté que les élèves utilisaient principalement des stratégies de savoir ou des stratégies mixtes, mais finalement peu de stratégies de repli (stratégies B). Cela ne signifie pas pour autant qu'elles ne se manifestent pas ; en particulier nous avons observé que les élèves les mobilisaient pour répondre à des QCM de résolution de problèmes complexes (Sayac & Grapin, 2014a). Pour des QCM portant sur des transformations d'écritures, les processus de réponse des élèves reposent la plupart du temps sur des connaissances mathématiques, parfois erronées. Même si la formulation des questions peut être différente de celle usuellement proposée en classe, il semblerait que les élèves

reconnaissent des tâches qui leur sont habituelles et cherchent à mobiliser leurs connaissances ; ce qui n'est pas toujours le cas, notamment lorsque la tâche est plus complexe.

Si globalement l'écart de score entre QCM et QROC n'est pas toujours significatif, nous avons pu néanmoins constater, en étudiant les réponses produites par les élèves individuellement que les connaissances mobilisées par l'un ou l'autre des formats de questions n'étaient pas identiques : ce qui remet en cause la validité « psycho-didactique » de la question, aussi bien pour les QCM évaluant des connaissances sur les entiers que sur les décimaux.

Dans cette recherche, au-delà de l'étude sur la validité du format QCM en évaluation, nous avons également souhaité rendre compte des difficultés que rencontrent les élèves français dans l'écriture des nombres à la fin de l'école primaire. Ce travail spécifique et local, puisque limité à un petit nombre d'items, autour de la validité d'un dispositif d'évaluation permet plus largement de poser la question des contenus d'évaluation et des choix que l'on peut être amené à faire pour essayer de repérer, sur des tâches précises, ce que les élèves savent ou ne savent pas faire. Ce n'est qu'une facette de la validité didactique à laquelle nous nous sommes intéressées ici à travers l'activité de l'élève ; elle est complémentaire à une étude plus globale du contenu de l'évaluation, prenant par exemple en compte, la représentativité des tâches relatives au domaine évalué.

Au-delà des résultats produits par ces tests à grande échelle pour les institutions et par ces expérimentations pour les chercheurs, on peut également imaginer l'intérêt qu'ils pourraient présenter pour l'enseignant. En effet, l'utilisation en classe par l'enseignant de tels indicateurs nous paraît être une voie intéressante pour lui permettre de mieux appréhender le niveau de ses élèves et réguler son enseignement. Il conviendrait donc de diffuser plus largement les résultats de telles enquêtes, mais en veillant à accompagner leur diffusion d'un temps de formation<sup>8</sup> pour permettre aux enseignants de les comprendre, dans la réalité interprétative que nous avons essayé de dévoiler.

Pour les élèves, nous rejoignons Pluvinaud (1979) qui affirme que « pointer cette erreur et faire remarquer son caractère attractif à un élève qui l'a commise est un service à lui rendre ». Ainsi, il pourrait être intéressant de confronter les élèves aux erreurs fréquentes repérées par les didacticiens

pour mieux déconstruire les conceptions qui les sous-tendent. Cette confrontation pourrait se réaliser à travers l'usage fréquent et régulier de QCM qui servirait alors d'évaluation diagnostique ou formative pour l'enseignant tout en étant très attractif pour les élèves (Abrahamson, 2006). Cette proposition rejoint le principe des boîtiers de vote électronique (également appelés clickers), utilisés dans certaines classes (Draper, 2005), et actuellement étudiés avec beaucoup d'intérêt aussi bien par les chercheurs (Foster & Miller, 2009; Marquet, 2011; Faillet, Marquet & Rinaudo, 2013) que par les institutions.

Nous voyons donc que l'étude des bilans évaluatifs dans une approche didactique permet de réaliser ce qui est réellement évalué du point de vue des connaissances en jeu, mais aussi de mieux comprendre comment les élèves s'emparent des questions posées selon les différents formats adoptés. Au-delà de ces constats, nous pensons que l'évaluation des apprentissages mathématiques des élèves doit faire l'objet d'une formation spécifique pour permettre aux professeurs de mieux exploiter les différentes modalités d'évaluation qui se trouvent à leur disposition (ce qui inclut les QCM) en fonction de leurs objectifs d'enseignement et ainsi en permettre un usage adapté selon les visées diagnostique, formative ou finale des évaluations qu'ils sont régulièrement amenés à concevoir dans leur classe.

## NOTES

1. CEDRE : cycle des évaluations disciplinaires réalisées sur échantillon. Ces évaluations sont réalisées par le ministère de l'Éducation nationale français tous les 6 ans, suivant différentes disciplines, depuis 2000. En, 2008 et 2014, elles ont porté sur les mathématiques.

2. Par opposition au QCM, nous qualifions d'« ouverte » une question qui n'a pas de proposition de réponse laissée au choix de l'élève.

3. Terme utilisé pour qualifier les choix de réponses d'un QCM autres que la (ou les) bonne(s) réponse(s).

4. **Modèle 1** : quand l'étudiant « sait », il choisit la réponse correcte et quand il ne « sait pas », il choisit au hasard parmi les réponses proposées.

**Modèle 2** : commence comme le Modèle 1, mais au lieu de répondre au hasard quand il « ne sait pas », l'étudiant commence par éliminer les solutions qu'il sait être fausses et choisit au hasard parmi celles qui restent.

**Modèle 3** : l'étudiant commence par ranger les solutions possibles par ordre de plausibilité décroissante et, si la consigne l'oblige à ne fournir qu'une d'entre elles, alors il choisit celle dont la probabilité (subjective) est la plus élevée (à ses yeux).

5. Définis par Leclercq et Poumay (2004, p. 3), les degrés de certitude sont « l'ensemble des jugements, des analyses, des régulations, conscientes ou non, (mais qu'il importe de rendre explicites, observables et conscientes) effectués par l'apprenant sur ses propres performances » ; nous avons proposé une échelle de quatre degrés de certitude, adaptée à des élèves de fin d'école : pas sûr du tout (1), pas très sûr (2), sûr (3), sûr et certain (4).

6. La répartition des réponses aux différentes QROC figure en annexe 2.

7. Rappelons que ces stratégies peuvent s'appuyer sur des savoirs corrects ou erronés (considérer par exemple que  $a/b = a/b$ ).

8. À ce titre, le dispositif PACEM proposé par Chesné (2014) est très instructif.



## RÉFÉRENCES

- Abrahamson, L. (2006). A brief history of networked classrooms : Effects, cases, pedagogy, and implications. Dans D. A. Banks (Ed.), *Audience response systems in higher education* (p. 1-25). Hershey, PA : Information Science Publishing.
- Adda, J. (1976). Difficultés liées à la présentation des questions en mathématiques. *Educational studies in mathematics*, 7, 3-22.
- Attali, Y., & Bar-Hillel, M. (2003). Guess Where : The Position of Correct Answers in Multiple-Choice Test Items as a Psychometric Variable. *Journal of Educational Measurement*, 40(2), 109-128.
- Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of Free-Response and Multiple-Choice Items. *Journal of Educational Measurement*, 22(1), 77-92.
- Bosch, M., & Gascon, J. (2005). La praxéologie comme unité d'analyse des processus didactiques. Dans A. Mercier & C. Margolinas (Eds.), *Balises pour la didactique des mathématiques* (p. 197-122). Grenoble : La Pensée Sauvage.
- Braibant, J.-M. (2015) Comment aider et accompagner les enseignants universitaires à améliorer la qualité des examens par QCM sur base des analyses d'items? *Actes du 27<sup>e</sup> Colloque de l'ADMEE Europe*. Liège.
- Bridgeman, B. (1992). A Comparison of Quantitative Questions in Open-Ended and Multiple-Choice Formats. *Journal of Educational Measurement*, 29(3), 253-271.
- Chesn , J.-F. (2014). *D'une  valuation   l'autre : des acquis des  l ves sur les nombres en sixi me   l' laboration et   l'analyse d'une formation d'enseignants centr e sur le calcul mental* (Th se de doctorat, universit  Paris-Diderot, Paris).
- Chevallard, Y., & Feldmann, S. (1986). *Pour une analyse didactique de l' valuation*. IREM d'Aix-Marseille.
- CHOPPIN, B. (1975). Guessing the answer on objective tests. *British journal of educational psychology*, 45, 206-213.
- COMITI, C., & NEYRET, R. (1979).   propos des probl mes rencontr s lors de l'enseignement des d cimaux en classe de CM. *Grand N*, 18, 5-20.
- Douady, R., & Perrin-Glorian, M.-J. (1986). *Liaison  cole – coll ge : Nombres d cimaux*. Brochure n  62. Paris : IREM de Paris 7.
- Draper, S. (2005). *Using EVS at Glasgow University*. Rep r    : [<http://evs.psy.gla.ac.uk/local.html>] consult  le 21/11/2017.
- Duval, R. (1996). Quel cognitif retenir en didactique des math matiques? *Recherches en Didactique Des Math matiques*, 16(3), 349-382.
- Duval, R. (2006). Transformations de repr sentations s miotiques et d marches de pens e en math matiques. *Actes du XXXII  colloque COPIRELEM* (p. 67–91). Strasbourg : Irem de Strasbourg.
- Duval, R., & Pluvineau, F. (1977). D marches individuelles de r ponse en math matique. *Educational studies in mathematics*, 8(1), 51-116.
- Faillet, V., Marquet, P., & Rinaudo, J.-L. (2013). L' l ve invisible : recherche sur l'utilisation des boitiers de vote au lyc e. *Sticef*, 20, 2013. Rep r  sur le site : [<http://sticef.univ-lemans.fr/>].
- Foster, D., & Miller, H. (2009). A new format for multiple-choice testing : discrete-option multiple-choice. Results from early studies. *Psychology science quarterly*, 51(4), 355-369.
- Forsyth, R. A., & Spratt, K. F. (1980). Measuring problem solving ability in mathematics with multiple-choice items : the effect of item format on selected item and test characteristics, *Journal of Educational Measurement*, 17(1), 31-43.
- Frary, R. B. (1985). Multiple-choice versus free-response : a simulation study, *Journal of Educational Measurement*, 22(1), 21-31.
- Gilles, J.-L. (1996). Utilisation des degr s de certitude et normes de r alisme en situation d'examen et d'auto-estimation   FA.P.S.E. – ULG. *Colloque de l'ADMEE-EUROPE « Dix ann es de travaux de recherche en  valuation »*.
- Ginsburg, H. P., Lopez, L. S., Mukhopadhyay, S., Yamamoto, T., Willis, M., & Kelly, M. S. (1992). Assessing Understanding of Arithmetic. Dans R. Lesh & S. J. Lamon (Eds.), *Assessment of Authentic Performance in School Mathematics* (p. 265-289). Washington : AAAS Press.
- Grapin, N. (2015). * tude de la validit  de dispositifs d' valuation et conception d'un mod le d'analyse multidimensionnelle des connaissances des  l ves de fin d' cole* (Th se de doctorat, universit  Paris Diderot. Paris).
- Gr goire, G., & Laveault, D. (2014). *Introduction aux th ories des tests en sciences humaines*. Bruxelles : De Boeck.
- Grugeon-Allys, B. (1997). Conception et exploitation d'une structure d'analyse multidimensionnelle en alg bre  l mentaire. *Recherches en Didactique des Math matiques*, 17(2), 167-210.
- Grugeon-Allys, B., Chenevotot-Quentin, E., Pilet, J., & Delozanne, E. (2012). Diagnostic et parcours diff renci s d'apprentissage en alg bre  l mentaire. *Recherches en Didactique des Math matiques, Hors-s rie*, 133-158.
- Grugeon-Allys, B., & Grapin, N. (2015). Validit  d'une  valuation externe : compl mentarit  entre une approche didactique et psychom trique. *Actes du s minaire national de didactique des math matiques*.
- Joffe, L. S. (1992). The English experience of a national curriculum and assessments. Dans G. C. Leder (Ed.), *Assessment and Learning of Mathematics* (p. 190-222). Hawthorn, Victoria : Australian Council for Educational Research.
- Katz, I., Bennett, R. E., & Berger, A. (2000). Effects of response Format on difficulty of SAT mathematics Items : it's not the strategy. *Journal of Educational Measurement*, 37(1), 39-57.



- Leclercq, D. (1986). *La conception des questions à choix multiple*. Bruxelles : Labor.
- Leclercq, D. (1987). *Qualité des questions et signification des scores*. Bruxelles : Labor.
- Leclercq, D. (2006). L'évolution des QCM. Dans G. Figari & L. Mottier-Lopez (Éd.), *Recherche sur l'évaluation en éducation* (p. 139-146). Paris : L'Harmattan.
- Leclercq, D., & Poumay, M. (2004). Une définition opérationnelle de la métacognition et ses mises en œuvre. *Communication présentée à la 21ème conférence internationale de l'AIPU*, Marrakech.
- Lukhele, R., Thissen, D., & Wainer, H. (1994). On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests. *Journal of Educational Measurement*, 31(3), 234-250.
- Marquet, P. (2011). e-Learning et conflit instrumental. *Recherche et formation*, 68, 31-46.
- Maury, S. (1985). Influence de la question dans une épreuve relative à la notion d'indépendance. *Educational studies in mathematics*, 16, 283-301.
- MEN (2015). *Note d'information de la DEPP, 18 mai 2015 – CEDRE 2014 : mathématiques en fin d'école primaire : les élèves qui arrivent au collège ont des niveaux très hétérogènes*, DEPP. France.
- Mounier, E. (2010). *Une analyse de l'enseignement de la numération au CP. Vers de nouvelles pistes* (Thèse de doctorat, université Paris Diderot. Paris).
- Pilet, J. (2012). *Parcours d'enseignement différencié appuyés sur un diagnostic en algèbre élémentaire à la fin de la scolarité obligatoire : modélisation, implémentation dans une plateforme en ligne et évaluation* (Thèse de doctorat, université Paris-Diderot, Paris).
- Pluvinaige, F. (1979). Loto-questionnaires (pour l'évaluation et l'auto-contrôle en mathématiques). *Educational studies in mathematics*, 10, 443-485
- RODITI, E. (2007). La comparaison des nombres décimaux, conception et expérimentation d'une aide aux élèves en difficulté. *Annales de didactique et de sciences cognitives*, 12, 55-81.
- Rodriguez, M. C. (2003). Construct Equivalence of Multiple-Choice and Constructed-Response Items : A Random Effects Synthesis of Correlations, *Journal of Educational Measurement*, 40(2), 163-184.
- Sayac, N., & Grapin, N. (2014a). Évaluer les capacités des élèves à résoudre des problèmes dans le cas d'une évaluation externe en France : les spécificités de la forme QCM. *Éducation et Francophonie*, XLII(2), 64-83.
- Sayac, N., & Grapin, N. (2014b). Évaluer par QCM en fin d'école : stratégies et degrés de certitude. *Annales de didactique et de sciences cognitives*, 19, 169-199.
- Sayac, N., & Grapin, N. (2015). Évaluation externe et didactique des mathématiques : un regard croisé nécessaire et constructif. *Recherches en Didactique des Mathématiques*, 35(1), 101-126.
- Tate, R. (2000). Performance of a Proposed Method for the Linking of Mixed Format Tests With Constructed Response and Multiple Choice Items, *Journal of Educational Measurement*, 37(4), 329-346.
- Vantourout, M., & Goasdoué, R. (2014). Approches et validité psycho-didactique des évaluations, *Éducation et Formations*, e-302, 139-156.
- Vantourout, M., & Maury, S. (2016). Impact des modalités des questions sur les performances des évalués : étude empirique en mathématiques en fin de collège. *Actes du 28<sup>e</sup> Colloque de l'ADMEE Europ*. Lisbonne.
- Vergnaud, G. (1981). Quelques orientations théoriques et méthodologiques des recherches françaises en didactique des mathématiques. *Recherches en Didactique des Mathématiques*, 2(2), 215-230.
- Yackel, E., Cobb, P., & Wood, T. (1992). *Instructional development and assessment from a socio-constructivist perspective*. Dans G. C. Leder (Ed.), *Assessment and Learning of Mathematics* (p. 63-82). Hawthorn, Victoria : Australian Council for Educational Research.

## ANNEXE 1

## Stratégies A : stratégies de savoirs

A1 : l'élève effectue la tâche demandée mentalement ou explicitement puis trouve, parmi les propositions, celle qui correspond à la réponse trouvée.

A2 : l'élève reconnaît d'emblée la « bonne » réponse parmi celles proposées (connaissance intériorisée).

A3 : l'élève applique une règle simple intériorisée, correcte ou non (théorème en actes).

A4 : l'élève teste les propositions de réponse une à une jusqu'à trouver celle qui convient.

## Stratégies B : stratégies de substitution ou de repli

B1 : l'élève répond au hasard.

B2 : l'élève passe en revue superficiellement toutes les propositions, puis choisit celle qui lui paraît la plus vraisemblable.

B3 : l'élève ne sait pas expliquer sa procédure.

B4 : l'élève combine les nombres en présence de manière à trouver, parmi les choix possibles, une solution.

## Stratégies C : stratégies mixtes

C1 : l'élève commence par s'engager dans une procédure de résolution, mais sans aller jusqu'au bout (à la différence de A1) ; il utilise ensuite les différentes propositions de réponses pour conclure (en choisissant celle qui lui paraît la plus vraisemblable).

C2 : l'élève élimine les propositions qui paraissent fausses, puis déduit de celle(s) qui reste(nt), la bonne réponse.

## ANNEXE 2

## Répartition des 195 réponses par item : QCM 1 et QCM 2

QCM 1a : Deux-millions-deux-cent-vingt-cinq-mille-six s'écrit :	Nombre d'élèves	Pourcentage
<input type="checkbox"/> 2 225 600	15	7,7 %
<input type="checkbox"/> 2 522 600	8	4,1 %
<input type="checkbox"/> 2 225 006	172	88,2 %
<input type="checkbox"/> 2 225 060	0	0 %

QCM 2	6 chiffres	7 chiffres	8 chiffres	9 chiffres
QCM 2a : Sept-millions-cinq-cent-cinq-mille	21 (10,8 %)	152 (77,9 %)	12 (6,2 %)	10 (5,1 %)
QCM 2b : Vingt-cinq-millions-cinq-cent-mille	25 (12,8 %)	19 (9,7 %)	144 (73,8 %)	7 (3,6 %)
QCM 2c : Cinq-cent-mille	180 (92,3 %)	8 (4,1 %)	5 (2,6 %)	2 (1 %)
QCM 2d : Sept-cent-millions-cinq-cent-mille	18 (9,2 %)	14 (7,2 %)	12 (6,2 %)	151 (77,4 %)

## Répartition des 195 réponses par item : questions ouvertes

## QROC 1a : Écrire en chiffres « cinq-millions-cinq-cent-cinquante-deux-mille-six »

Réponses produites	Nombre d'élèves	Pourcentage
5 552 006	137	70,3 %
5 552 600	2	1,1 %
Absence de zéros (55 526)	7	3,8 %
Absence d'un zéro (5 552 06 ou 5 55026)	5	2,7 %
Erreur liée au rang des millions (5 000 552 006)	1	0,5 %
Erreur sur un des chiffres (4 552 006 ou 5 542 006...) autre que celui des unités	9	4,9 %
Écriture de 0 au lieu de 6 au chiffre des unités (5 552 000)	5	2,7 %
Transcription écrit – oral (5 000 000 552 000 6000 ou 5 000 000 550 2006 ou 5 550 2006)	3	1,6 %
Autres erreurs liées à la position des chiffres dans l'écriture	7	3,8 %
Autres erreurs (plusieurs chiffres erronés et/ou mauvaise position)	16	8,6 %
Non réponse	3	1,6 %

**QROC 1b : Écrire en chiffres :**  
**huit-millions-six-cent-six-mille**

Réponses produites	Nombre d'élèves	Pourcentage
8806000	138	74,6 %
Absence de zéros (866)	1	0,5 %
Absence de zéros terminaux (8606)	15	8,1 %
Absence d'un ou deux zéro terminaux (8 606 0 ou 8606 00)	3	1,6 %
Erreur liée au rang des millions (8 000 606 000 ou 8606000000)	5	2,7 %
Erreur sur un des chiffres (7 606 000 ou 8 506 000...)	1	0,5 %
Transcription écrit – oral (8 000 000 600 6 000 ou 8000000 606000)	1	0,5 %
Autres erreurs liées à la position des chiffres dans l'écriture	12	6,1 %
Autres erreurs (plusieurs chiffres erronés et/ou mauvaise position dans l'écriture)	13	6,6 %
Non réponse	6	3,1 %

**QROC 2 : Quel est le nombre de chiffres nécessaires pour écrire chacun des nombres ?**

Réponses produites	Nombre de chiffres								
	10 ≤	9	8	7	6	5	≥ 4	Écriture du nombre	Non réponse
QROC 2a : Trente-quatre-millions-cinq-cent-mille	4	6	134	3	9	16	12	9	2
	2 %	3,1 %	68,7 %	1,5 %	4,6 %	8,2 %	6,2 %	4,6 %	1 %
QROC 2b : Neuf-cent-mille	0	2	0	3	149	3	21	9	8
	0 %	1 %	0 %	1,5 %	76,4 %	1,5 %	10,8 %	4,6 %	4,1 %
QROC 2c : Huit-cent-millions-deux-cent-mille	5	138	4	6	16	4	8	9	5
	2,5 %	70,8 %	2,1 %	3,1 %	8,2 %	2,1 %	4,1 %	4,6 %	2,5 %

**QROC 3 : Donner une écriture à virgule de la fraction  $\frac{95}{10}$**

Réponses produites	Nombre d'élèves	Pourcentage
9,5	153	78,5 %
0,95	12	6,2 %
95,10	13	6,7 %
950	0	0,0 %
9,5/10	4	2,1 %
Autre réponse	8	4,1 %
Non réponse	5	2,6 %

**QROC 4 : Écrire 149,7 sous forme d'une fraction**

Réponses produites	Nombre d'élèves	Pourcentage
1 497/10	101	51,8 %
149 + 7 octobre	3	1,5 %
149/7	16	8,2 %
1 497/100	2	1,0 %
1 497/1 000	16	8,2 %
149/10	7	3,6 %
149,7/10	4	2,1 %
Autre réponse	25	12,8 %
Non réponse	21	10,8 %

## ANNEXE 3

## Répartition des stratégies selon les QCM

QCM 1a : Choisir l'écriture en chiffres de « Deux-millions-deux-cent-vingt-cinq-mille-six »	Nombre d'élèves	Pourcentage
Stratégies de savoir	189	96,9 %
Stratégies de substitution ou de repli	1	0,5 %
Stratégies mixtes	5	2,6 %

QCM 2 : Choisir le nombre de chiffres qui composent l'écriture de quatre nombres entiers	Nombre d'élèves	Pourcentage
Stratégies de savoir	178	91,3 %
Stratégies de substitution ou de repli	10	5,1 %
Stratégies mixtes	7	3,6 %

QCM 3 : Choisir une écriture à virgule de $\frac{62}{10}$ ?	Nombre d'élèves	Pourcentage
Stratégies de savoir	182	93,3 %
Stratégies de substitution ou de repli	13	6,7 %
Stratégies mixtes	0	0,0 %

QCM 4 : Choisir une écriture fractionnaire de 237,8	Nombre d'élèves	Pourcentage
Stratégies de savoir	175	89,7 %
Stratégies de substitution ou de repli	6	3,1 %
Stratégies mixtes	14	7,2 %